

統制語による検索の未来

さけみ かよ
酒見 佳世

(メディアセンター本部)

1. はじめに

普段、何気なく使う言葉に、統一性があるかどうかを意識することはなく、また意識する必要もない。例えば、あなたの職業は？と聞かれてなんと答えるだろうか。司書、図書館員、ライブラリアン、メディアセンター職員。毎回違う言葉で答えても、何の問題もない。しかし、データベースを検索するということになる、状況は一変する。

2. 統制語とは何か

統制語とは、その意味範囲や使い方が規制された言葉である。統制語を使用すれば、同義語、類義語、同形異義語、ことばの意味の曖昧さなどによって生じる検索漏れやノイズを回避することが出来る¹⁾。この統制語を一覧表にしたものがシソーラスや件名標目表である。

3. カード目録からオンライン目録へ

1980年代ごろから、図書館目録はカードからオンラインに移行し始め²⁾、90年代半ばごろからはウェブOPACに移行してきた。カード時代には書名の前方一致か著者名目録、主題検索をする際は、分類目録か少数の図書館で作成される件名目録と、資料にアクセスする手段は限られていた。オンライン目録への移行により、カード目録ではできなかった項目による検索が可能になり、さらにブール演算子によって、検索結果の絞り込みや拡大ができるようになった³⁾。その結果、主題からアプローチする場合、カード目録時代には直接書架に向かっていた利用者が、目録でキーワード検索を行うことが多くなった⁴⁾。Drabenstottは1991年の研究レビューにおいて、OPACは利用者に非常に好まれており、OPACで多くの主題検索が行われているとまとめている。しかし、同時に件名を使用したOPACの主題検索には問題があり、改善の必要があるとも指摘している⁵⁾。Larson(1991)の研究では、問題の多い件名検索の利用が次第に減少していく一方で、タイトル

キーワード検索は増加していくという傾向が示されている⁶⁾。タイトルキーワード検索は、不完全ではあるが、OPACでの自然語からの検索を実現するものである。この自然語とは、統制語と対立する概念である。

4. 目録中のコントロール(統制)

目録中の書誌は記述の部分と、標目の二つの部分に分けられる。記述部分には現物からとられた書誌事項が記載され、標目部分には、記述部分からアクセスポイントとして選択され、統制された事項が記述される。書誌中で統制されるのは、著者名(団体名、会議名を含む)、タイトル、シリーズ名、件名である。この作業によって、検索の一貫性が保障される。しかし、この作業を行うためには、名称典拠の維持が必要である。この典拠の必要性はインターネットの世界が広がるにつれて再認識されるようになってきている⁷⁾。日本においては国立国会図書館により「全日本典拠総合データベース」(仮称)が検討されているが⁸⁾、現状では、総合的な名称典拠を持ち、ウェブで公開している図書館は存在しない。現在の慶應の目録においても典拠そのものは維持されていない。

5. 主題をあらわす統制語：件名

このように書誌中には統制された語がいくつか存在していることになる。この中で、主題検索のために付与されるのが件名である。図書館で、主題検索のために用意されているツールは、分類と件名であるが、日本の大学図書館で主に使用されてきたのは分類である。この理由としては、書架配置を主題別(分類番号順)にしていたこと、大学図書館で使えるレベルの件名標目表が存在しないこと、きちんと維持され公開されている名称典拠が存在しなかったことが考えられる。また、オンライン目録に移行してからは、主題検索はタイトルキーワード検索によって代替されてきたと考えられる。

しかしながら、件名検索は日本の大学図書館のウェブOPACの60%以上で可能となっており、少なくとも図書館では、主題からの検索項目として主要な位置づけを与えられているといえる⁹⁾。一方、過去のいくつかの利用者調査^{10)~12)}から、利用者側にも、主題検索のニーズ自体は存在すると考えられる。この場合、問題となるのは目録の提供側と、利用者側のギャップである。

6. KOSMOS IIでの件名利用

実際に件名がOPACでどの程度使用されているのかを知るため、KOSMOS IIのトラザクシヨログの分析¹¹⁾を行った。調査対象としたのは2004年10月の1ヶ月間のログで、全423,804件である。表1は主な検索項目の利用件数と全検索数に占める割合を示したものである。なお、この件数には掛け合わせ検索の数は含まれていない。

表1から分かるように、現状のKOSMOS IIで、件名・分類はあまり利用されていない。件名の利用率は総検索数の2.1%、分類は0.1%である。この結果から、多くの主題検索はタイトルキーワード(「書名・誌名中の語」あるいは「全て」)から行われているとみられる。キーワードからの検索には長所、短所がいくつかある。短所のうち、問題となるのはテキストに含まれてはいるが、タイトルに表現されない情報が検索できないこと、また包括的なリンクが欠如することである¹³⁾。例えばGerhan (1989)は、タイトルキーワードのみからの検索では目録中の全ての関連図書の55%しか検索できなかったと述べている¹⁴⁾。

表1. 主な検索項目の利用件数と全検索数に占める割合 (2004年10月)

検索項目	検索数	全検索数に占める割合
全て	20,146	4.8%
書名・誌名中の語	200,337	47.3%
正確な書名・誌名	38,629	9.1%
著者名	32,555	7.7%
出版社	842	0.2%
件名	9,107	2.1%
分類	512	0.1%
全検索	423,804	100.0%

表2. ゼロヒットの割合

検索項目	ゼロヒットの割合	ゼロヒット件数
全て	28%	5,577
書名・誌名中の語	26%	51,853
正確な書名・誌名	4%	1,403
著者名	32%	10,441
出版社	28%	237
件名	14%	1,368
分類	18%	95
全検索	26%	111,091

検索でヒットがなかったものの割合(ゼロヒット率)を表2に示した。ゼロヒットの件数は全体で111,091件(総検索数の26%)であったが、これに比べて「件名」のゼロヒット率は14%と低かった。これは件名利用の75%がリンクあるいは索引語検索を使用しているためであると考えられる。

7. 利用者と検索語

日本ではもともと件名目録が普及しておらず、また最近ではウェブの利用がOPACの利用にも影響を及ぼしているということもあり¹⁵⁾、利用者が統制語を意識して、OPACに検索語を入力するという状況はより考えにくいものとなっている。こうした中、2003年7月に発表された、BatesによるLCのアクションプランに対する報告書において、目録作成者ではなく、利用者が検索に利用した語彙を集めて辞書を作成し、その語彙と主題検索語(件名)をリンク付けすることで利用者を正しい検索語に導こうという提案がなされている¹⁶⁾。

8. 件名の利用法

Batesの提案するような仕組みをすぐ実現するのは難しいと思われるが、現状で考えられる件名の利用について3点ほどあげておく。

1点目は、NIIが提供している米国議会図書館件名標目表(LCSH)¹⁷⁾の翻訳版を使用することである。基本件名標目表(BSH)、国立国会図書館件名標目表(NDLSH)と比べて、統制語の絶対量を増やすことが可能になる¹⁸⁾ので、結果として検索語との一致率が

増加すると考えられる。この件名をキーワード検索に利用することで、利用者が思いついた語から検索しても、件名が付与されていない場合よりもヒット率は上昇するはずである。

2点目は、件名を二次的な検索に使用することである。これにより、網羅的な主題検索を実現することができる。この場合、最初はまず思いついた語で検索し、検索結果の中から適切なレコードを選び、その書誌データに付与された件名のリンクをクリックすることになる。

3点目は、ブラウジング検索の利用である。KOSMOS IIでは索引語一覧という形で、ブラウジング検索を提供している(図1参照)。この検索の利点は件名が事前結合の統制語であることから生まれるものである。主標目とそれに続く細目で出来上がっている件名を一覧することによって、利用者は自分では思いつかなかった主題の多くの面を認識し、自分にとって必要な資料を選び出すことが可能になると考えられる¹⁶⁾。

No.	索引語	件数
1	Iran War, 2003	49
2	Iran War, 2003 — Accidents	2
3	Iran War, 2003 — Biographies	1
4	Iran War, 2003 — Causes — Causes	1
5	Iran War, 2003 — Causes — Third States	1
6	Iran War, 2003 — Causes	11
7	Iran War, 2003 — Command operations	1
8	Iran War, 2003 — Consequences	2
9	Iran War, 2003 — Diplomatic history	1
10	Iran War, 2003 — Great Britain	1
11	Iran War, 2003 — Influence	2
12	Iran War, 2003 — Jewish literature	1
13	Iran War, 2003 — Law and literature	1
14	Iran War, 2003 — Mass media and the war	11
15	Iran War, 2003 — Military operations	1
16	Iran War, 2003 — Moral and ethical aspects	2
17	Iran War, 2003 — Moral and ethical aspects — Third States	1
18	Iran War, 2003 — Occasional literature	1
19	Iran War, 2003 — Peace	1
20	Iran War, 2003 — Personal narratives	1

図1. KOSMOS II 索引語一覧の検索例

9. おわりに

米国議会図書館の目録政策・支援室長の Tillet は目録の作業の定型的、日常的な部分の多くが機械化され、カタログはより知的な視点が必要とする語彙の統制や主題付与、分類作業に従事することが可

能になっていると述べている¹⁷⁾。目録は一朝一夕に出来上がるものではなく、日々少しずつ作られるものである。すぐに役立つものもあるが、長期的に続けていくことで効果を発揮するようになるものもある。件名付与もその一つであろう。

OPAC で、また OPAC につながる全てのデータベースで、電子情報源で、主題から網羅的にもれなく検索できること。これが統制語によって実現される未来である。

参考文献

- 1) 図書館情報学ハンドブック編集委員会. “5. 情報検索”. 図書館情報学ハンドブック. 第2版. 東京, 丸善, 1999, p. 535-648.
- 2) 北克一, 芝勝徳, 志保田務. 書誌情報の標準化と OPAC—1980年代以降の動向と発展. 図書館界. vol. 45, No. 1, 1993. p. 123-141.
- 3) 平輪麻里子. ネットワーク検索時代におけるキーワードの役割. 医学図書館. vol. 45, no. 2, 1998, p. 222-227.
- 4) 吉田憲一. 主題検索と OPAC—司書過程の学生への目録利用調査から. 図書館学会年報. vol.40(2), 1994, p.71-84.
- 5) Drabentstott, K. M. Online catalog user needs and behavior: Think tank on the present and future of the online catalog. Proceedings, Chicago, American Library Association, 1991. RASD Occasional Papers, no. 9.
- 6) Larson, Ray R. The decline of subject searching: long-term trends and patterns of index use in an online catalog. JASIS. vol. 42 (3), 1991, p. 197-215.
- 7) 鈴木智之. パーチャル国際典拠ファイル: その試みと可能性. カレントアウェアネス. no. 280, 2004, p. 2-3.
- 8) 渡邊隆弘. セマンティックウェブと図書館. カレントアウェアネス. no. 281, 2004, p. 9-12.
- 9) 3) と同じ
- 10) 佐川祐子. 公立図書館における OPAC の導入: 東京都杉並区立中央図書館の利用記録の分析を通して. 現代の図書館. vol. 28, no. 2, 1990, p. 66-75.
- 11) 「主題情報の組織化とユーザー・インターフェイス」研究班. 利用者の検索行動と主題情報: 国立国会図書館における OPAC モニター調査を中心に. 図書館研究シリーズ. 32, 1992, p. 179-269.
- 12) 入戸奈実. 大学図書館における OPAC の利用者の探索行動: 学生を対象としたインタビュー調査. Library and

- Information Science. vol. 37, 1998, p. 35-53.
- 13) Dubois, C. P. R. Free text versus controlled vocabulary: a reassessment. Online Review. vol. 11, no. 10, 1987, p. 243-253.
- 14) Gerhan, David R. LCSH in vivo: subject searching performance and strategy in the OPAC era. Journal of Academic Librarianship. vol. 15, 1989, p. 83-89.
- 15) 種市淳子, 逸村裕. 短期大学生の情報探索行動の分析. 三田図書館・情報学会研究大会発表論文集. 2005, p. 37-40.
- 16) Mann, Thomas. Why LC Subject Headings are more important than ever. American Libraries. vol. 34, no. 9, 2003, p. 52-54.
- 17) 上保佳穂. デジタル環境における目録作成: バーバラ・B・ティレット米国議会図書館目録政策・支援室長講演会報告. 国立国会図書館月報. no. 496, 2002, p. 20-25.
- とシステムとの応答を記録したもので, 利用者の検索意図や満足度を明らかにすることはできないが, 利用者の目録検索過程とその結果についての詳細な情報が得られる。
- 2) ベイツレポートに関しては, 以下に詳しく紹介されている。
橋詰秋子. 米国にみる「新しい図書館目録」とその可能性: ベイツレポートを中心に. 現代の図書館. vol. 41, no. 4, 2003, p. 222-230.
- 3) LCSH については, 以下に詳しく紹介されている。
鹿島みづき. LCSH とメタデータ標準的テーマスキーマの応用が意図するもの. 大学図書館研究. no. 71, 2004, p. 1-10.
- 4) 各件名標目表の最新版と件数は以下の通り。
LCSH: 第 28 版 (2005), 標目数 約 28 万件
NDLSH: 第 5 版 (1991), 標目数 17,133 件
BSH: 第 4 版 (1999), 標目数 7,847 件

注

- 1) トランザクションログとは, オンライン目録での利用者